

# Global oceanographic variational data assimilation of in-situ observations and space-borne altimeter data for reanalysis applications

Andrea Storto<sup>1</sup>, Srdjan Dobricic<sup>2</sup>, Simona Masina<sup>1,2</sup> and Pierluigi Di Pietro<sup>1</sup>

<sup>1</sup>Istituto Nazionale di Geofisica e Vulcanologia, Bologna, Italy, storto@bo.ingv.it

<sup>2</sup>Centro Euro-Mediterraneo per i Cambiamenti Climatici, Bologna, Italy

## 1. Introduction

The study of global climatological trends requires the accurate analysis of the surface and sub-surface ocean state. In the last two decades, altimetric satellite missions have been launched with the aim of monitoring the sea-level height variability, in time and space. This information may in turn be used, within data assimilation systems, for adjusting the column-integrated density fields in synergy with in-situ observations. The impact of the Sea Level Anomaly (SLA) data has been recently proved positive in many regional and global data assimilation system. However, gaining a positive impact from altimetric data needs i) the establishment of a correct strategy for updating temperature and salinity fields accordingly; ii) the correct assessment of the Mean Dynamic Topography to add to the anomaly data; iii) the consistency between the scales represented by the SLA data and those resolved by the ocean model.

At the National Institute for Geophysics and Volcanology (INGV) and the Euro-Mediterranean Centre for Climate Change (CMCC), the former reduced-rank Optimal Interpolation (OI) analysis system (Bellucci et al., 2007) was used to produce ocean reanalysis for the last four decades. It has recently been replaced with a three-dimensional variational data assimilation system, which uses a First Guess at Appropriate Time (FGAT) algorithm. The 3DVAR/FGAT formulation is adapted from the one operationally used for producing daily analysis in the Mediterranean basin (Dobricic et al., 2008), and is able to successfully assimilate satellite sea-level anomaly observations.

## 2. Formulation of the data assimilation scheme

The assimilation system is based on a 3DVAR/FGAT formulation, which consists of minimising a functional  $J$  given by

$$J = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \frac{1}{2}(\mathbf{y} - \mathbf{H}(\mathbf{x} - \mathbf{x}^b) - \mathcal{H}(\mathbf{x}^b))^T \mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}(\mathbf{x} - \mathbf{x}^b) - \mathcal{H}(\mathbf{x}^b))$$

where  $\mathbf{x}$  is the analysis at the minimum of  $J$ ,  $\mathbf{x}^b$  is the background, which is a prior estimate of the state of the ocean,  $\mathbf{y}$  is the vector of the observations,  $\mathcal{H}$  is the fully non-linear observation operator which projects the state of the ocean onto the space of the observations,  $\mathbf{H}$  is the tangent-linear version of the observation operator and  $\mathbf{B}$  and  $\mathbf{R}$  are the covariance matrices of the background and observational errors, respectively. In the formulation of the previous equation, the fully non-linear observation operator is used only once for computing the initial departures using the background fields closer to observation time (the so-called First Guess at Appropriate Time). The tangent-linear model is used for updating the cost function at each iteration according to the new model state, while the adjoint model for mapping the new observation departures back into the model space for the gradient computation. Their linearization is performed around the background fields closer to observation time. In our formulation,

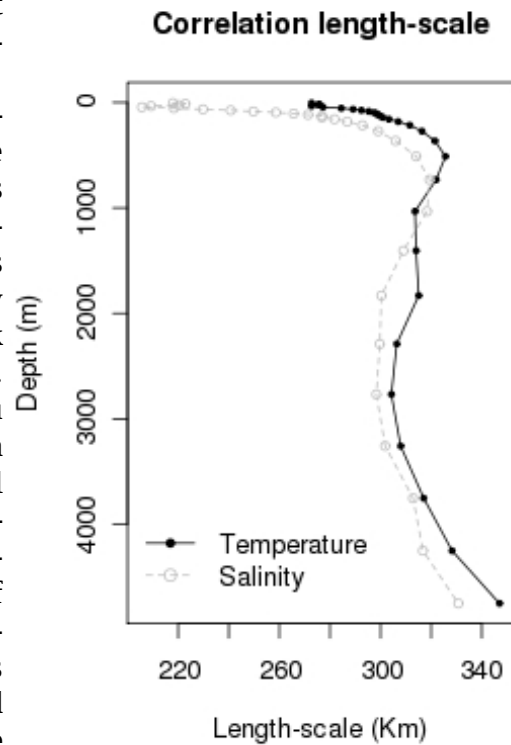
only temperature and salinity are corrected after a 3DVAR/FGAT assimilation step, so that the model space is composed of the pair of (T, S) in the three-dimensional ocean grid.

The formulation of the background term of the cost function follows Dobricic and Pinardi (2008), who firstly use a “change-of-variable” operator to precondition its minimization, and then decompose the root-square inverse of the background error covariance matrix into two linear operators, which account for vertical covariances and horizontal correlations, respectively. In particular, vertical covariances are represented by seasonal column-independent multivariate 10-mode EOFs of temperature and salinity in the full model grid, deducted from the model climatology. Horizontal correlations are obtained by means of an application of four iterations of a first-order recursive filter (Hayden and Purser, 1995). The vertically-varying correlation length-scale has been computed by simulating the errors as differences of 5- and 15-day forecasts valid at the same nominal time (the so-called “NMC method”). In Figure 1 the length-scale is shown as a function of depth for temperature and salinity, the latter exhibiting a shorter correlation within the first 1000 m.

The configuration of the system includes i) the 3DVAR/FGAT assimilation step, whose assimilation time-window is ten days long, and whose nominal analysis time is centred within it and ii) the ocean forecast model, which is run for fifteen days in order to ensure the availability of fields for the next FGAT assimilation step. The nominal analysis time is then located on the tenth day with respect to the previous forecast model step. The ocean forecast model is the Ocean Parallèle model (OPA) in its version 8.2 (Madec et al., 1998).

As surface boundary conditions, the ocean model uses sea surface temperature (SST), heat fluxes and wind stress from the ERA-40 reanalysis dataset.

The set of in-situ observations consists of vertical profiles of temperature from Expandable bathythermographs (XBT), buoys, sea stations (TESAC), Argo profiles (only from late 90s onwards) and salinity profiles from Argo, buoys and sea-stations. Data have been provided by the Met Office Hadley Centre in the framework of the EU-funded project ENSEMBLES (EN3). EN3 is a collection of quality-checked in-situ data with a time-dependent fall rate correction for XBT (Wijffels et al., 2008). An additional background quality check, which rejects observations whose departures from the model-equivalents is larger than three times the sum of background and observational errors is also performed on top of the minimisation. This avoids spurious inconsistencies between the model and the observations. Finally, a thinning procedure retains only the closest observation to analysis time if multiple reports from the same platform or buoy are found in close proximity.



**Figure 1: Globally-averaged correlation length-scale for use within the recursive filter.**

### 3. Use of SLA data

The observed sea level anomaly  $\zeta^o$  is directly related to the sea surface height through the

relation

$$\zeta^o = \eta - \eta_m + \varepsilon$$

where  $\eta_m$  is the Mean Dynamic Topography (MDT),  $\eta$  is the sea surface height (SSH) and  $\varepsilon$  accounts for the errors of the SLA observation, sea surface height and MDT. According to the 3DVAR/FGAT formulation, the SLA contribution to the observational term of the cost function is

$$\mathbf{y} - \mathbf{H}(\mathbf{x} - \mathbf{x}^b) - \mathcal{H}(\mathbf{x}^b) = \zeta^o - \mathbf{H}(\mathbf{x} - \mathbf{x}^b) - (\eta^f - \eta_m)$$

where  $\eta^f$  is the sea surface height prognosed by the ocean model, which includes the computation of the surface freshwater budget. To compute the tangent-linear increments of the SLA model-equivalents within the minimisation, we use a “local hydrostatic adjustments” (LHA) scheme, which is based on the vertical integration of density increments:

$$\rho_0 \delta \eta + \int_{-H}^0 \delta b(T, S) dz = \delta p_b$$

where  $\delta \eta$  is the height increment given by the vertical integration of the buoyancy  $\delta b$  ( $b = \rho' / \rho_0$ ) between the bottom level  $H$  and the surface, function of the temperature  $T$  and the salinity  $S$ . The previous equation is solved by assuming the existence of a “level of no-motion” (1500 m), when applicable, corresponding to the depth  $H$ , where horizontal velocities are practically zero (which implies that the buoyancy increment is zero,  $\delta p_b = 0$ ). In practice, our scheme splits the observation departure in thermo- and halo-steric contributions over the water column. The accuracy of such a scheme with respect to the prognostic formulation of the SSH in the ocean model is found to be of about 1-2 cm far from the Equator, reasonably below the nominal instrumental precision, and less than 5 cm close to Equator ( $\pm 5$  degrees of latitude).

The experiments presented later on covered the period between October 1992 and January 2006, during which all the available altimetric data from TP, ERS-1 and -2, GFO, Jason-1 and ENVISAT have been assimilated. Data are provided by AVISO, after the usual geophysical removals (tropospheric, ionospheric, electromagnetic, tidal and inverse barometer effects) and after a multi-satellite cross-calibration for eliminating residual orbit errors and large-scale biases (Le Traon et al., 1998). To filter out high frequencies which are not resolved by the model, we apply a low-pass Lanczos filter with a latitudinally-varying cut-off wavelength, consistent with the model and the data resolution. The filter decreases the global RMSE against observations of about 2-3 mm. Like the in-situ observations, sea level anomalies observations are quality-checked against model-equivalents values. Finally, a satellite-dependent thinning procedure retains only one observation on a box of size 150% the model resolution. This ensures the consistency between the assimilated SLA dataset with the scale resolved by the ocean model and decreases the spatial correlation of the SLA observational errors, removing also the correlation introduced by the Lanczos filter.

Sea level anomaly data need to have a mean dynamic topography to be added to for comparison with the model sea surface height. Purely observational method, like that described in Rio and Fernandez (2004, hereafter RIO04) are based on space-borne gravimetric data and eventually adjusted by means of in-situ observations. They may not represent the mean dynamic topography seen by the ocean model, whose geoid, sea-ice exchanges and run-off representation is much more simplified than in real-life. We have therefore firstly derived the MDT as the model mean SSH by using weighed analysis and short-range forecasts initialised by assimilating in-situ data only, and then adjusting this MDT through assimilation diagnostics. Following Dobricic (2005), we assume that the observation minus guess bias for the SLA data

is dominated by the bias in the MDT; consequently, we may use the observation minus guess data to correct the MDT, by means of an univariate OI. The use of such a derived MDT, however, is equivalent to assimilate the sea-level height variability rather than the sea-level height itself, which would require a full consistency between the observed MDT and the model MDT.

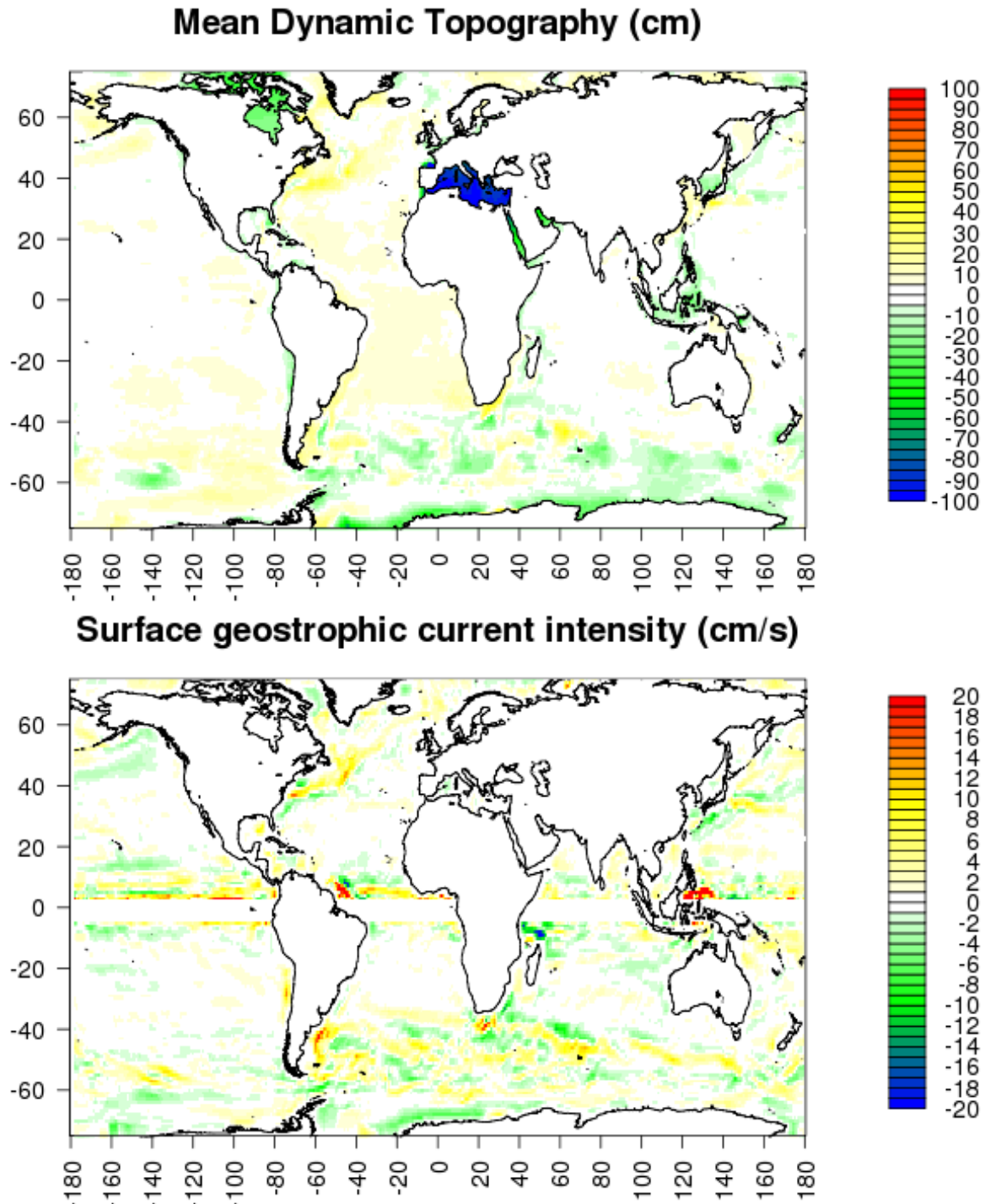
In Figure 2 we show a comparison between the RIO04 MDT and that computed with our two-step procedure, in terms of absolute unbiased difference (top panel) and difference of surface geostrophic current intensities (bottom panel), deducted from the MDT maps. Besides the large bias in the Mediterranean basin, we found that main differences are located in areas of large eddy activity and cannot be simplified to a global offset. The differences are particularly large and noisy in the mid- and high-latitude southern hemisphere. Note that in those areas, further to have a strong mesoscale activity, the lack of a dense network of in-situ observations decreases the accuracy of the RIO04 MDT. Surface geostrophic currents deducted from the RIO04 MDT result differently located in the Gulf Stream Region and in correspondence of the Caribbean, the Falkland, the Agulhas and Equatorial Counter currents, and in the Kuroshio current area.

#### **4. Selected Results**

The 3D-Var system (with SLA data assimilation) and the formerly used reduced-rank Optimal Interpolation analysis system show in general very comparable results in terms of mean temperature and temperature anomaly trends, except a localised increase of mean temperature in the thermocline in the Equatorial Atlantic, in accordance with the World Ocean Atlas 2005 (WOA05) climatology, and a slight shift of the larger temperature gradient in the Gulf Stream region towards the North-American coast.

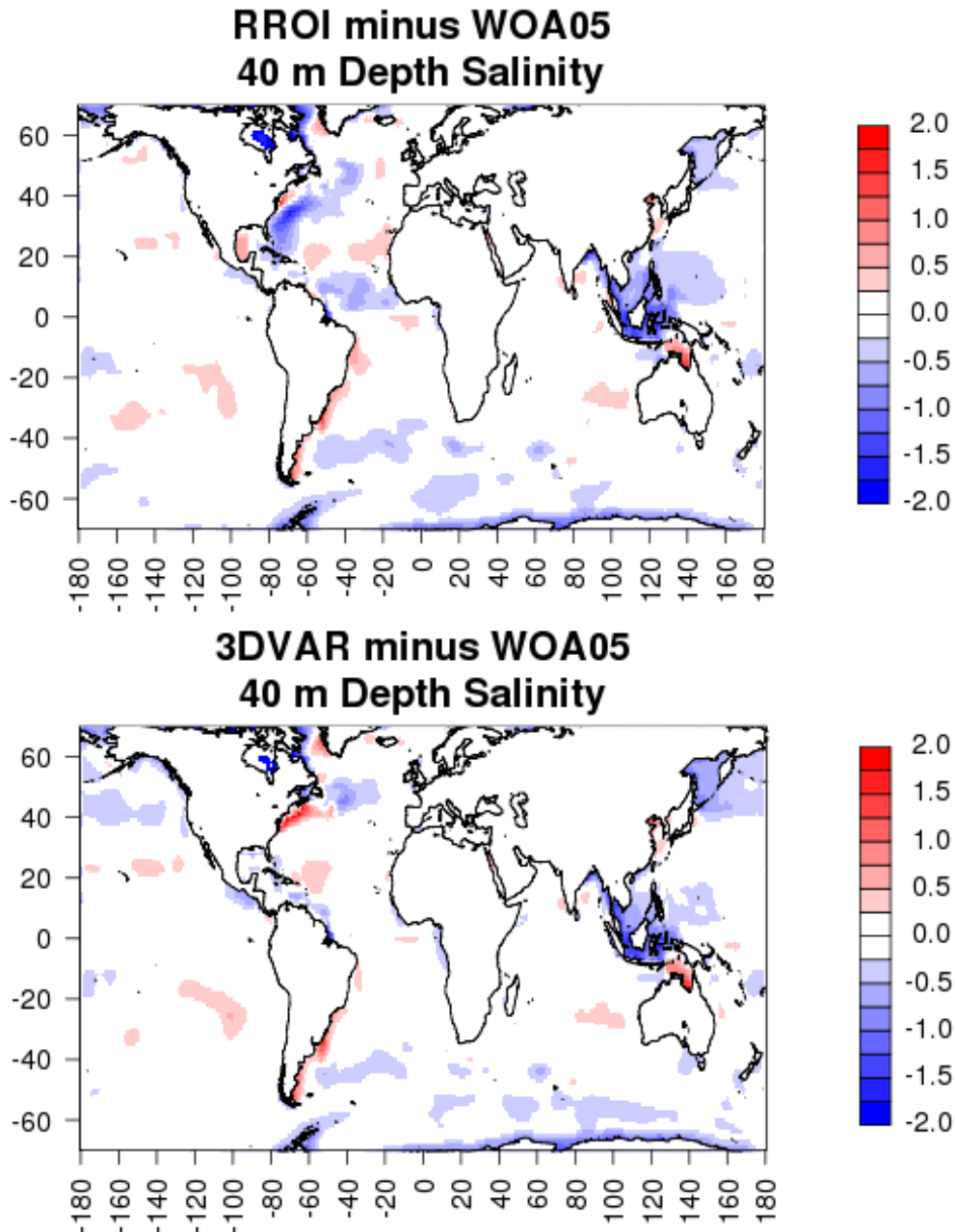
The impact of the data assimilation system is seen to be very important for improving the near-surface salinity fields throughout the global ocean. The differences with the climatology decrease, especially in the North and Equatorial Atlantic, where the climatological salinity exhibits an increase of about 1 PSU in the western part of the Gulf Stream Region, and a smaller decrease on the European side. In the Equatorial Atlantic, the new assimilation system causes a salinity increase in the thermocline and a decrease below. Figure 3 shows the difference of the 12-year model climatology with the WOA05 climatology for the 40m deep salinity for both the assimilation schemes. Most of differences between the two climatologies are located in the Atlantic Ocean.

We found that the RMSE of the model sea-level height against the SLA observations decreases significantly when sea-level anomaly observations are assimilated. This is particularly evident in the North Atlantic and North Pacific regions where the RMSE decrease is of about 2 to 4 cm, and even more in the southern hemisphere, where it reaches 8 cm. Note that as our SLA assimilation scheme corrects only near- and sub-surface fields of temperature and salinity, the result about the better fit of the model SSH to the SLA observations indicates that satellite altimetric data successfully correct the temperature and salinity on the vertical. On the contrary, the use of the RIO04 MDT increases the RMSE in most of regions of about 2-3 cm with respect to the OI MDT. This confirms that the impact of the MDT on the assimilation system is large, and care must be taken in its representation.



**Figure 2: Top panel: contour map of difference between the RIO04 MDT and the MDT calculated through optimal interpolation; bottom panel: contour map of difference of surface geostrophic current intensity.**

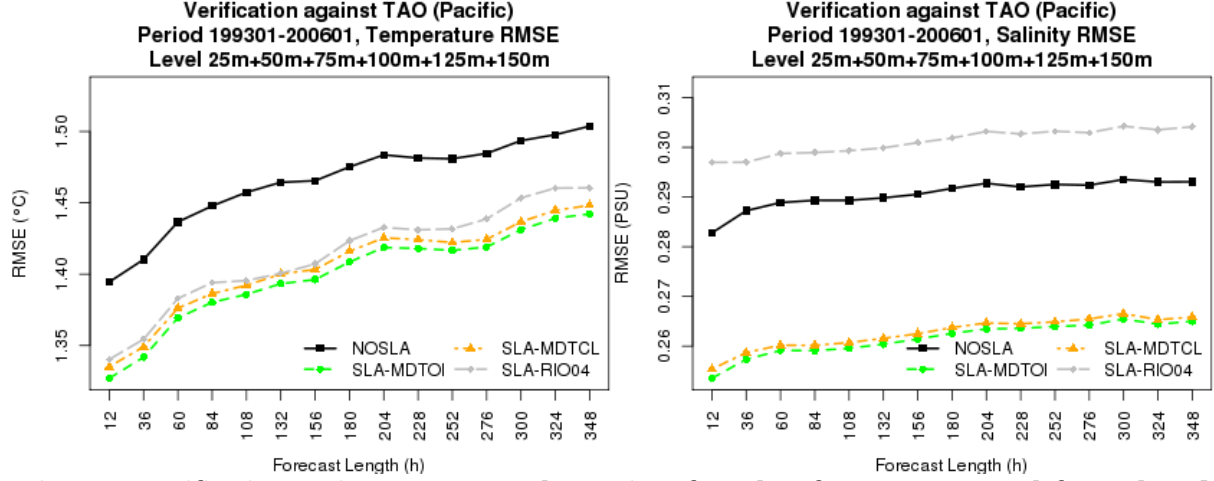
Verification against independent observations has been conducted for the whole experimental period. In particular, Figure 4 shows the root mean square error of the forecasts against the Tropical Atmosphere Ocean project (TAO) in the Equatorial Pacific area, for the thermocline region. The impact of the SLA observations on subsurface temperature skill scores is clearly of benefit for the assimilation and forecast system, whatever MDT is used, though the MDT calculated through OI of assimilation statistics performs slightly better than the other MDT at



**Figure 3: Difference with the World Ocean Atlas 2005 climatology: reduced-rank OI (RROI, top panel) and 3D-Var analysis system (3DVAR, bottom panel).**

all forecast lengths. For the salinity skill scores, we found that the use of the RIO04 MDT is detrimental, while the use of a model-based MDT induces positive impact of SLA data. The impact on deeper levels is much smaller. Likewise for the surface temperature and salinity skill scores, due to the model surface forcing. Verification skill scores have also been computed in the Tropical Atlantic (PIRATA array, only temperature) and Tropical Pacific (RAMA array), showing qualitatively similar results to the TAO ones, with the exception that in the Atlantic the temperature scores are significantly worse when the RIO04 MDT is used.

Regarding the impact of SLA data on the 12-year climatology, we found an increased near- and sub-surface salinity in the whole Pacific, especially at mid-latitudes, and, in all the oceans, in correspondence of the southern hemisphere sub-tropical gyre. On the contrary, the



**Figure 4: Verification against TAO array observations for subsurface temperature (left panel) and salinity (right panel) as a function of forecast range. NOSLA is the experiment without assimilation of sea-level anomaly observations; SLA-MDTOI, SLA-MDTCL, SLA-RIO04 refer to the experiments with assimilation of sea-level anomaly observations, using as MDT that from OI, from model climatology and from RIO04, respectively.**

mean salinity decreases in the Indian Ocean and, on the average, in the Northern Atlantic. The impact on the temperature is more noisy, besides a clear warming effect of SLA observations in correspondence of the Kuroshio region and the North-Atlantic gyre and the Equatorial Atlantic, while exhibits a cooling effect in the tropical Pacific.

## 5. Conclusions

A global three-dimensional variational assimilation system has been implemented with the aim of producing ocean reanalysis for the last three decades. The comparison with the formerly used reduced-rank Optimal Interpolation analysis system proves that the new assimilation scheme is particularly of benefit for the subsurface salinity fields, especially in the Atlantic Ocean.

In the new global 3D-Var analysis scheme, we implemented the assimilation of sea-level anomaly observations from altimetric satellites. SLA data are assimilated through a local hydrostatic adjustment scheme. Vertical profiles of salinity and temperature are simultaneously corrected, according to the density increment corresponding to the sea-level anomaly. We used a low-pass Lanczos filter to filter out the SLA data high-frequency signals, not resolved by the coarser resolution model. Results show a positive impact of the SLA data, and the large importance of the choice of the MDT, especially visible in the salinity verification skill scores. Mean dynamic topography was taken from model climatology and adjusted to account for MDT bias through an optimal interpolation scheme. This is equivalent to assimilate sea-level variability instead of sea-level height itself, which would need the model mean SSH to be consistent with an observed MDT. In the future, this might be achieved, for instance, by forcing the model mean SSH to the observed MDT and using more sophisticated sea-ice and run-off schemes.

## Acknowledgements

This work has been partly supported by the INGV study programme “Programma Internazionale di Studi Avanzati sull'Ambiente e sul Clima”, funded by the “Fondazione Cassa di Risparmio di Bologna”. The authors want to thank the AVISO team for the support in the use of SLA data, as well as Simon Good (UK MetOffice) for the support in the use of the EN3 dataset.

## References

- Bellucci, A., S. Masina, P. Di Pietro, and A. Navarra, 2007: Using temperature–salinity relations in a Global Ocean implementation of a multivariate data assimilation scheme. *Mon. Wea. Rev.* , **135** , 3785–3807.
- Cooper, M. and K. Haines, 1996: Altimetric assimilation with water property conservation. *J. Geophys. Res.* , **101** , 1059–1077.
- Dobricic, S., 2005: New mean dynamic topography of the Mediterranean calculated from assimilation system diagnostics. *Geophys. Res. Lett.* , **32** , L11606.
- Dobricic, S. and N. Pinardi, 2008: An oceanographic three-dimensional assimilation scheme. *Ocean Modelling* , **22** , 89–105.
- Dobricic, S., N. Pinardi, M. Adani, C. Fratianni, A. Bonazzi, and V. Fernandez, 2008: Daily oceanographic analyses by the Mediterranean basin scale assimilation system. *Ocean Science* , **3** , 149–157.
- Hayden, C. M. and R. J. Purser, 1995: Recursive filter objective analysis of meteorological fields: Applications to NESDIS operational processing. *J. Appl. Meteor.* , **34** , 3–15.
- Le Traon, P., F. Nadal, and N. Ducet, 1998: An improved mapping method of multi-satellite altimeter data. *J. Atmos. Ocean. Technol.* , **15** , 522–534.
- Madec, G., P. Delecluse, M. Imbard, and C. Levy, 1998: OPA 8.1 Ocean General Circulation Model reference manual. Note du pole de modélisation no 11, Institut Pierre-Simon Laplace. Pp. 91.
- Rio, M. H. and F. Hernandez, 2004: A mean dynamic topography computed over the world ocean from altimetry, in-situ measurements and a geoid model. *J. Geophys. Res.* , **109** , C12032.
- Roullet, G. and G. Madec, 2000: Salt conservation, free surface, and varying levels: a new formulation for ocean general circulation models. *J. Geophys. Res.* , **105** , 23927–23942.
- Vidard, A., M. Balmaseda, and D. Anderson, 2009: Assimilation of altimeter data in the ECMWF Ocean Analysis System 3. *Mon. Wea. Rev.* , **137** , 1393–1408.
- Wijffels, S. E., J. Willis, C. M. Domingues, P. Barker, N. J. White, A. Gronell, K. Ridgway, and C. J. A., 2008: Changing expendable bathythermograph fall rates and their impact on estimates of thermosteric sea level rise. *J. Climate* , **21** , 5657–5672.